

The Relative Influence of Research on Class-Size Policy

JAMES S. KIM

Social science research suggests that reducing class size has its largest effects on the achievement of minority and inner-city children during the first year of formal schooling.¹ Despite scholarly disagreements about the implications of specific studies on class size, economists generally agree that targeted class-size policies rest on stronger evidence than untargeted policies. For example, economist Eric Hanushek contends that “surely class-size reductions are beneficial in specific circumstances—for specific groups of students, subject matters, and teachers.”² Similarly, economist Alan Krueger notes that the “effect sizes found in the STAR experiment and much of the literature are greater for minority and disadvantaged students than for other students [and] economic considerations suggest that resources would be optimally allocated if they were targeted toward those who benefit the most from smaller classes.”³ However, a number of state legislatures have enacted untargeted and expensive policies to reduce class sizes in all schools, among all subgroups of students, and beyond the early elementary grades. Therefore, the central tension between research and policy in the class-size debate is this: research seems to support targeted class-size policies most strongly, but targeted policies are the exception rather than the norm in the policy arena. As a result, some social scientists have criticized across-the-board class-size reductions as prohibitively expensive and scientifically indefensible.⁴

The Politics of Class Size

In this paper, I argue that the influence of social science research on class-size policy depends partly on political context. Instead of simply asking

Table 1. Classifying and Explaining the Politics of Class-Size Reduction Policies

<i>Benefits</i>	<i>Distributed costs</i>	<i>Concentrated costs</i>
<i>Distributed benefits</i>	Majoritarian politics (Untargeted class-size reduction)	Entrepreneurial politics
<i>Concentrated benefits</i>	Client politics (Targeted class-size reduction)	Interest group politics

whether research affects policy, it is equally important to examine how the politics of the class-size debate shape the many uses of research in the policy arena. Social scientists cannot control the uses of research, but they can encourage evaluation of policies in terms of their effects on educational outcomes. Social scientists possess methodological tools rather than policy prescriptions, and research is a valuable resource because it helps clarify the goals of public policies without dictating the outcome of democratic debate.

Political scientist James Q. Wilson provides a way of classifying the politics surrounding policy issues such as class-size reduction. Wilson's framework can help identify some conditions under which research may influence public policy. As shown in table 1, untargeted class-size reduction represents a popular education policy that distributes benefits and costs among a large number of people. Public opinion polls suggest that a majority of voters—including teachers, superintendents, parents, and taxpayers—support reducing class sizes in order to improve their local schools.⁵ When legislators cast class-size reduction as “majoritarian issue,” the proposal appeals to large blocs of voters and it is generally easy to form a majority coalition supporting it. A recent example of an untargeted class-size policy is the 2002 Florida class-size amendment, which was supported by a motley collection of voters—senior citizens, union leaders, parents, and educators—who approved a multibillion-dollar policy that mandates a cap on class size in elementary, middle, and secondary schools by 2010. The Florida policy, as well as similar untargeted class-size policies in California and Nevada, follows the prediction advanced by Wilson's framework: “If a new policy is adopted and people become convinced that the promised benefits are real and worth the cost, debate ends and the program will not only continue but will also grow rapidly in size.”⁶ Widespread political support is sufficient to legitimize an untargeted class-size reduction policy. Therefore there is little incentive for legislators to plan for evaluation of a popular policy whose central goal is to advance political interests rather than to achieve preset educational goals.

Table 1 also shows that targeted class-size reduction policies distribute costs and concentrate benefits among “clients” of government. According to Wilson,

when “benefits are concentrated, the group that is to receive those benefits has an incentive to organize and work to get them.”⁷ Evidence plays a more vital role in helping a policy survive and grow when it concentrates benefits on a subgroup of clients—in this case, young children who neither vote nor pay taxes. Because perceived benefits (both monetary and nonmonetary) are not distributed among a large number of individuals and interest groups, such policies do not automatically command a wide base of political support. Instead, targeted policies must rely on alternative forms of support, such as empirical evidence that demonstrates the efficacy of an intervention or normative claims about the value of improving the achievement of minority and low-income children. When class-size policy is cast as a client-based issue, there is usually a stronger incentive for state legislatures to commission research, to apply findings to class-size policy, and to draw on empirical evidence to defend and maintain the policy. In addition, there is an incentive to appeal to values and to argue that young children from minority and low-income families are legitimate and deserving clients of government. The values underpinning compensatory education programs remain strong in American society and may deepen support for targeted class-size reductions in schools and classrooms with large numbers of disadvantaged children.⁸ Targeted class-size policies illustrate cases in which research can play a direct and meaningful role in the political arena by informing debate and guiding policy decisions.

This chapter examines the influence of research on class-size policy in three sections, organized chronologically. First, it reviews empirical research on the relationship between class size and student achievement. Second, it describes the role that research played in shaping class-size policies in Tennessee and Indiana in 1989 and in Wisconsin and California in 1996. Each pair of case studies illustrates substantial differences in the way state legislators framed the class-size debate and used research in determining the scope and size of their respective policies. Third, it discusses several lessons bearing on the relationship between research and policy as applied to class size in particular and education reform in general.

Research on Class Size: Evidence from the 1960s and 1970s

Research on class size evolved over the twentieth century from complex statistical analyses of national surveys to innovative meta-analyses of primary studies and ultimately, in Tennessee, to the gold standard in social science—the randomized field trial. The largest national survey to shed light on the

relationship between educational inputs, such as class size, and student achievement was the 1966 Equality of Educational Opportunity survey, more famously known as the Coleman Report. Surprisingly, sociologist James Coleman and his colleagues found that “some facilities measures, such as the pupil/teacher ratio in instruction . . . showed a consistent lack of relation to achievement among all groups under all conditions.”⁹ Similarly, in one reanalysis of the Coleman data, sociologist Christopher Jencks concluded that “while reductions in class size can often be justified in terms of teachers’ sanity, pleasant classroom atmosphere, and other advantages, they are hard to justify in terms of test scores.”¹⁰ However, since those findings were based on analyses of observational data, they were merely suggestive and tentative. The dearth of large experiments further impeded understanding of whether reduced class sizes *caused* improvements in student achievement. Summarizing the state of knowledge on class size in the early 1970s, John Gilbert and Frederick Mosteller noted that “after a half a century of tightly controlled studies of optimum class size, we have made practically no progress. . . . the studies have been too small and specialized for their implications to have much chance of holding in new situations.”¹¹

Glass and Smith’s Meta-Analysis of Class Size

By the late 1970s, research findings on the effects of class size had been accumulating for nearly a century and the time was ripe for a “summing up” of the empirical evidence. In 1978, Gene Glass and Mary Lee Smith conducted a summary of class-size research, using an innovative but controversial technique called meta-analysis—a study of studies.¹² In their review, Glass and Smith found seventy-seven studies on the relationship between class size and student achievement, conducted between 1900 and 1978. Using meta-analysis to combine results across studies, Glass and Smith found that achievement increased “when class size is reduced below 20.”¹³ However, since the relationship between achievement and class size was not linear, achievement improved more dramatically when class size was reduced from twenty to ten students than from thirty to twenty students.¹⁴

The meta-analytic findings prompted debate among scholars. In one of the first criticisms of Glass and Smith’s meta-analysis, the Educational Research Services (ERS) faulted the meta-analysis for drawing unwarranted recommendations from poorly controlled studies because only fourteen of them were true experiments.¹⁵ The ERS study also showed that substantial achievement gains resulted only when class size was reduced to one to five students in small

tutorial sessions, a prohibitively costly and infeasible policy option for most public school districts. Similarly, Robert Slavin noted that one of the studies with the largest effect sizes in the Glass and Smith meta-analysis was on tennis instruction, in which the achievement outcome involved "rallying a tennis ball against the wall . . . as many times as possible in 30 sec."¹⁶ Another social scientist, Hans Eysenk, argued that meta-analysis was "an exercise in megasilliness," and he suggested that the adage "Garbage in equals garbage out" applied to the Glass and Smith synthesis of class size.¹⁷ Glass retorted that the ERS report was "a dog's breakfast of pleonasm, confusion, and obfuscation."¹⁸ Regardless of one's perspective, the debate between Glass and his critics represented what Harry Aaron called "self-cancelling" research—conflicting evidence from poorly designed research that paralyzed action.¹⁹ Given the lack of scientific consensus on the effects of class size, few states relied on the Glass and Smith meta-analysis to enact class-size reduction initiatives in the early 1980s.²⁰

Tennessee's Project STAR

During the mid-1980s, social scientists played a key role in encouraging the Tennessee state legislature to evaluate the effects of small classes on student achievement rather than implement a statewide reduction in class sizes. In 1984, as part of the broader political debate over education reforms initiated by Governor Lamar Alexander, educators and teacher unions began to push for legislation that would reduce the pupil-teacher ratio to 20:1. However, several policymakers "were not convinced that the state of research warranted the high-cost of reducing class sizes statewide."²¹ The uncertainty about the effects of small class size on achievement encouraged two policy entrepreneurs to call for a more targeted and limited policy. In particular, Steve Cobb, a state legislator trained as a social scientist, and Helen Bain, an educator with deep political connections, helped lawmakers focus the scope and size of the class-size policy. The class-size debate prompted Cobb to review the Glass and Smith meta-analysis, which suggested that class sizes had to be reduced to nearly fifteen students in order to improve student achievement. Therefore Cobb concluded that an across-the-board reduction would be expensive and ineffective and began to encourage his colleagues to support a controlled evaluation of a class-size reduction initiative. Bain, a former president of the National Education Association and a college professor, also argued that class size should target the early grades, when children first start school. Eventually, Cobb and Bain worked together to forge legislation, unanimously approved by

the Tennessee legislature, that authorized a demonstration project called Project STAR (Student/Teacher Achievement Ratio).²²

The STAR legislation is notable because it focused on specific educational goals, carefully defined the treatment under study, authorized a controlled investigation, and carved out prominent roles for social scientists. Project STAR was enacted in order "to study the effects of a reduced pupil-teacher ratio on the achievement of students in public school." In addition, each class in the demonstration project was "to have a maximum enrollment of seventeen." And the legislation called for "the identification of a control group of pupils in the same school system for the purposes of measuring differences in achievement."²³ The law also authorized the state department of education to invite researchers from four Tennessee universities to carry out the evaluation. Although the original STAR legislation did not require a randomized controlled trial, it gave researchers considerable freedom in designing the evaluation. To implement an experimental design, the researchers developed a protocol for the random assignment of students to small classes with thirteen to seventeen students, regular classes with twenty-two to twenty-five students, or regular classes with teacher aides.

During its first year, the Project STAR study, which has been celebrated as one of the greatest experiments in education, involved more than 6,000 students in seventy-nine elementary schools.²⁴ The STAR findings have been amply documented by researchers, but the most noteworthy findings were as follows: small class sizes of thirteen to seventeen students improved student achievement by approximately .20 of a standard deviation, or 4 percentile points; effect sizes were nearly twice as large for minority students as white students; test score gains were largest in kindergarten and first grade; and long-term effects persisted in a variety of academic outcomes measured in middle and high school.²⁵ Social scientists have conducted numerous secondary analyses using the STAR data, and policymakers have relied on the findings to justify class-size reduction policies.²⁶ In short, the study has been influential among scholars and politicians.

Tennessee's Project Challenge and Indiana's Prime Time Policy

When Project STAR ended in 1989, its findings were available to state policymakers interested in using research to develop class-size policy. How, then, did the research on class size influence state policy? Case studies of Tennessee and Indiana suggest that research played a very different role in shaping each state's policy.

From Tennessee's Project STAR to Project Challenge

In 1989, a newly elected Democratic governor, Ned McWherter, used the STAR results to craft a targeted class-size policy.²⁷ According to professor Charles Achilles, a principal investigator of the Tennessee study, "Ned looked at STAR and was most impressed with the results for poor and minority youngsters."²⁸ As a result, the governor "challenged" the poorest districts in Tennessee to cobble together federal Title I dollars and any other local revenue source to reduce class sizes from kindergarten through third grade, and he promised to match local efforts with discretionary funds from the state budget. Thus, Project Challenge distributed costs and burdens on taxpayers by using federal, state, and local money to reduce class sizes to an average of fifteen students from kindergarten through third grade in seventeen districts (ultimately sixteen districts by 1992–93) with the lowest per capita income.²⁹ In sum, Project Challenge cost nearly \$4.1 million to reduce class sizes in the poorest public school districts.

Research played a decisive role in legitimizing the Project Challenge policy, which was designed to concentrate educational benefits on minority and low-income children at risk of falling behind academically in the early grades. Ultimately, the sixteen Project Challenge districts involved a small proportion of Tennessee's 139 school districts and covered only 4 percent of classrooms in the state.³⁰ Although Project Challenge was not an experiment, the intervention seemed to raise the performance of the targeted districts on a standardized test of reading and mathematics from near the bottom to the median rank for all Tennessee districts.³¹ In the words of statistician Frederick Mosteller, Project Challenge shows how policymakers "can make use of a valuable intervention very selectively . . . by applying it to those who need it most."³²

Indiana's Prime Time Policy

In 1981, the Indiana legislature and then Governor Robert Orr enacted a state policy, dubbed Prime Time, to reduce pupil-teacher ratios in order to "improve basic skills in reading, writing, and arithmetic among Kindergarten, first, and second grades by reducing class size."³³ The Prime Time program—which began in 1981–82 as a small-scale pilot that provided \$300,000 to nine schools to reduce the student-teacher ratio to 14:1 in grades K-2—rapidly expanded into a statewide effort to reduce the pupil-teacher ratio in all districts by the end of the decade. The performance of Prime Time classes of 14 or fewer students was compared to regular class sizes that averaged nearly 23 stu-

dents. The results of the first evaluation revealed that positive effects were observed only in reading and only in first grade. Although these findings supported implementation of only a limited and targeted class-size reduction policy, they were used to support a major expansion of Prime Time.³⁴ By 1989–90, the Indiana legislature had committed more than \$65 million to fund class-size reductions in all 302 public school districts.³⁵

During the expansion of the Prime Time program in the 1980s, the state undertook a series of small studies that examined teaching and learning inside smaller classrooms. Early evaluations were based on nonexperimental studies, including ethnographies, classroom observations, and surveys of principals, teachers, and parents.³⁶ For example, the director of Prime Time noted that “evaluations of the first three years of the Prime Time program show happier, more productive teachers, more responsive students, and very satisfied parents” and that teachers had “more time for creative experiences, hands-on activities, drama, science exploration, and the use of learning centers.”³⁷ In a 1990 review of studies on Prime Time, however, one analyst noted that class-size reductions were “hastily implemented” and that “little has been done to objectively determine what effect Prime Time has had on educational achievement or to train teachers to take advantage of smaller classes.”³⁸

The most comprehensive evaluation of Prime Time is notable because it occurred nearly ten years after statewide implementation of the policy and ultimately supported a targeted reduction in class sizes. In 1999, the state superintendent of public instruction requested an independent evaluation of the Prime Time policy by researchers at Ball State University. According to the lead researcher, Daniel Lapsley, Prime Time was a popular policy and there was little pressure from politicians to evaluate it. Instead, it was the superintendent who wanted to obtain evidence of program effectiveness based on “an elaborate quasi-experimental study that longitudinally track[ed] students in Prime Time programs and compared them with students not in Prime Time programs.”³⁹ However, the evaluation followed a statewide class-size reduction program that had been in existence for nearly a decade, and the Ball State researchers highlighted the serious challenges and obstacles to performing an evaluation under such circumstances. The evaluation lacked sufficient controls to estimate the treatment effects, and the use of different versions of standardized tests in different schools precluded comparison based on a common metric.

Nonetheless, the evaluation reported findings consistent with those of previous research. The report’s central finding was that “small class size was not associated with better achievement, except for reading and composite achievement of minority pupils.”⁴⁰ This finding, the authors concluded, seemed to

indicate that "a differentiated Prime Time strategy might be most appropriate, one that seeks to reduce class size in lower SES (socio-economic status) schools."⁴¹ More precisely, these results supported a targeted class-size reduction policy for some students in some schools. Although some lawmakers have recently sought to target Prime Time funding as part of a categorical grant for at-risk students, it has been difficult to restrain a policy that provides additional money to virtually every district in the state. As one program administrator observed, the Prime Time program "is so much a part of the Indiana system [that] it would have been very difficult to completely take it away."⁴²

Wisconsin SAGE Program and the California Class-Size Reduction Policy

By 1996, the existing research on the effect of small class sizes on student outcomes was strengthened by the Project STAR experiment in Tennessee, which inspired both federal and state efforts to reduce class sizes. According to Marshall Smith, "nearly all of these initiatives are motivated, at least in part, by the findings of the Tennessee experiments."⁴³ In 1996, Wisconsin and California enacted class-size policies that represented two very different policy strategies. The centerpiece of the Wisconsin SAGE policy was a targeted reduction of class sizes to 15 or fewer students in kindergarten and first grade in twenty-one high-poverty districts. Given the small scale of the program, the legislature spent only \$4.2 million to fund SAGE schools in the 1996-97 school year. The California policy was an untargeted class-size reduction policy involving all California school districts that cost approximately \$1 billion in 1996-97, the first year of implementation. Although both policies have grown in size to include more students in more grades, the Wisconsin and California legislature used research in different ways to develop their policies.

Wisconsin SAGE Program

The Wisconsin SAGE program was conceived as a targeted class-size reduction program reserved primarily for lower-performing students in the early elementary grades who lived in high-poverty districts. In 1994, the state department of public instruction created the Urban Initiative Task Force, which included thirty-four members representing diverse political constituencies in Wisconsin. The primary goal of the task force was to discuss strategies for addressing the achievement gap and improving performance among low-

income children. Wisconsin policymakers therefore framed the problem as one of low achievement among a subgroup of students and recommended a targeted policy to address the issue. In 1996–97, the Wisconsin legislature funded a limited five-year pilot program that provided money to high-poverty schools to reduce class sizes to an average of 15 students in kindergarten and first grade, with additional grades to be added in 1997–98 (second grade) and 1998–99 (third grade).⁴⁴

Since the purpose of SAGE was to examine the effects of class-size reduction on children's academic outcomes, the Wisconsin legislature required an evaluation to determine whether the intervention accomplished its educational objective. The legislature authorized researchers at the University of Wisconsin–Milwaukee to conduct an evaluation comparing the academic performance of thirty SAGE schools and fourteen to seventeen comparison schools matched on student achievement and demographic characteristics.⁴⁵ The SAGE program was accompanied by a mandatory evaluation, and the quasi-experimental design was the next best alternative to a true experiment. It also was capable of addressing the relevant policy question—does class-size reduction in the early grades improve the cognitive skills of children in high-poverty elementary schools?

The year 1 (1996–97) and year 2 (1997–98) evaluation of the Wisconsin SAGE program produced findings that replicated the key findings from Tennessee's Project STAR. Similar to the schools in the Tennessee study, SAGE schools were required to reduce class sizes by having 1 teacher per 15 students or 30 students with 2 teachers in a single classroom. The sizable reduction in class size meant that pupil-teacher ratios were, on average, between 12:1 and 15:1 in SAGE classrooms and between 21:1 and 25:1 in comparison classrooms. Given the substantial contrast in the number of students in SAGE and control classrooms, Alex Molnar and his colleagues pointed out that "these differences were larger than class-size reductions in the Tennessee experiment and would be predicted to have measurable effects based on the Tennessee evidence."⁴⁶ In fact, the results of the early SAGE evaluations mirrored the STAR findings. In 1997–98, first-graders in SAGE classrooms scored approximately one-fifth of a standard deviation higher than the control group. The effects were largest for black children and for students attending the Milwaukee public schools, an urban district with a large percentage of low-income children.⁴⁷

The promising findings from the first- and second-year evaluation prompted the Wisconsin legislature to expand the SAGE program to include more than 500 schools by year 3 (2000–01). Critics of the rapid scale-up of SAGE have argued that class-size reduction is now too broad and expensive, that the bene-

fits will be diffused, and that the goal of assisting disadvantaged children will be undermined. According to the Wisconsin Policy Research Institute, the new and expanded version of the SAGE program "has lost its primary focus and, arguably, its initial justification."⁴⁸ The report concluded that a more accurate interpretation of existing evaluations would support a limited and cost-effective policy since "only African-American students in SAGE do better than African-American students in larger classes, and they do so only in first grade, and only in certain subjects and classrooms."⁴⁹ More recently, researchers at the University of Wisconsin–Madison conducted an analysis of long-term effects of the SAGE programs on kindergarten and first-grade students who were part of the original evaluation. The analysis confirmed significant program effects on standardized tests in grades 1 through 3 and replicated earlier analyses showing larger effects of SAGE on black students. However, there were no significant differences between SAGE and comparison students on a fourth-grade standardized test in reading comprehension, language arts, and mathematics.⁵⁰ Thus, two recent studies imply that a targeted class-size reduction policy would be more cost-effective than an untargeted policy in the early grades.

As with the evaluation of Tennessee's Project Challenge, evaluation of Wisconsin's SAGE program was planned before statewide implementation. That decision increased the odds that the results would be meaningful to researchers and policymakers. The availability of data from a well-designed quasi-experimental study enabled three independent organizations to evaluate SAGE in terms of its demonstrated impact on student learning—the initial goal of the program.⁵¹ The results did not prescribe a policy, but they helped focus policy on specific educational goals.

California's Class-Size Reduction Policy

In many ways, the 1996 class-size reduction policy in California was motivated by an educational crisis that demanded immediate political action. The crisis was prompted in 1994 by a combination of factors: California fourth-graders had tied for last place on the NAEP reading test; schools had an average elementary school class size of 29 students, which was the largest in the nation; and large achievement gaps persisted between white and minority students. In response to the perceived crisis in education, the California State Department of Education, led by then superintendent Delaine Eastin, convened the California Reading Task Force, whose goal was to identify possible solutions for improving reading among elementary school children. The task force identified small class sizes as a promising remedy for raising achievement in the ele-

mentary grades. In addition, California legislators began to scrutinize the results of the Tennessee STAR study, which showed significant benefits of reduced class sizes on the achievement of minority and inner-city children in the early grades. To obtain additional information on the effects of class-size reduction and develop different strategies for implementing such a policy, the state legislature requested an analysis by the California Research Bureau (CRB), a nonpartisan state research organization modeled after the federal government's Government Accountability Office.

In a June 1996 report entitled "Reducing Class Size: A Review of the Literature and Options for Consideration," David Illig, an analyst with the CRB, conducted a comprehensive review of the available research and recommended a range of policy options for lawmakers. The CRB report provided state legislators with a clear analysis of the research on class size and policy options.⁵² It noted that California's interest in class-size reduction "was motivated in part by the findings reported by the Tennessee Project STAR research team."⁵³ The report highlighted the findings of STAR, underscoring the conditions under which smaller classes improved achievement. Notably, the report discouraged an across-the-board reduction in class sizes. It also discounted the Glass and Smith findings in favor of the Tennessee STAR findings, used the STAR results to encourage a targeted class-size reduction policy in kindergarten and first grade and in high-poverty schools, and urged an independent evaluation of the state policy. Finally, the report raised questions about the generalizability of the Tennessee findings to California, which had a large percentage of English learners, limited space in overcrowded schools, and a shortage of credentialed teachers in urban schools.

The CRB report had very little impact on state policymakers. In July 1996, the Class-Size Reduction Program reduced class size to 20 students per teacher in kindergarten through third grade and was passed into law in one day.⁵⁴ The purpose of the state policy was "to increase student achievement, particularly in reading and mathematics, by decreasing the size of the K-3 classes to 20 or fewer students per certified teacher."⁵⁵ It was difficult to determine whether this goal was met because the rapid scale-up of the class-size policy made it virtually impossible to conduct a carefully controlled study. Evaluations funded after implementation provided weak and inconclusive evidence that class-size reduction improved achievement.⁵⁶ The authors of a final evaluation completed in 2001 concluded that class-size reduction "was nearly complete, and as a result we could not examine differences in SAT-9 scores between students who were and were not in reduced size classes."⁵⁷ The policy also had several negative and unintended consequences for urban districts and high-poverty

schools. California's class-size reduction initiative was "associated with declines in teacher qualifications and a more inequitable distribution of credentialed teachers."⁵⁸ In light of the unclear results from evaluations of the California class-size policy, researchers have encouraged more controlled experiments that would address a lingering question: Is class-size reduction a cost-effective strategy for improving student achievement, and if so, under what conditions?⁵⁹

Such questions assume that the California policy was designed primarily to produce measurable improvements in reading and mathematics achievement—an explicitly educational goal. In many ways, however, California's politicians viewed a massive, statewide effort to reduce class sizes as a response to an educational crisis—low NAEP scores in reading. Moreover, an untargeted policy satisfied the interests of a large number of groups. The perceived benefits were broadly distributed among state leaders wanting to demonstrate action in response to an educational crisis; parents, teachers, and union leaders wishing to reduce overcrowding in classrooms; and political leaders hoping to mollify various interest groups.⁶⁰

Sociologist Carol Weiss notes that in an atmosphere of crisis, research becomes ammunition to support predetermined positions "even if conclusions have to be ripped out of context."⁶¹ For California legislators, reducing class size in all schools seemed like the best strategy for improving academic achievement for all subgroups of students. Yet the Tennessee experiment's evidence was strongest for a class-size policy that targeted minority and inner-city children, and state policy analysts further underscored the cost-effectiveness of a targeted policy. However, it is unlikely that research will be called on to restrain the growth of a policy that distributes perceived benefits among voters in every California school district.

Discussion

The case studies suggest that the politics of the class-size debate often determine what, if any, effect research has on policymakers. Rather than asking why research does or does not influence policy, it seems equally important to consider how politics shapes the many uses of research in the policy arena. In doing so, I elaborate on five lessons that are relevant for the class-size policy in particular and education policy in general.

Lesson 1: When a policy enjoys a broad base of political support, there is virtually no incentive for state legislators to commission research on the pol-

icy or to evaluate it in terms of its demonstrated effect on student achievement. Since untargeted class-size reduction policies distribute benefits and costs, they inspire support among large numbers of voters and diverse constituencies. Popular policies can thrive in the absence of research evidence. The imperative to evaluate an untargeted class-size policy is weak because its primary goal is to advance political interests rather than to meet preset educational goals. Most state policies include statutory language stating that the goal of smaller class sizes is to improve reading and mathematics skills. Whether that goal is met or an evaluation is planned to measure performance is largely irrelevant. What matters most is whether the policy captures enough political support to survive. Votes, not evidence, are the touchstone for judging success. Therefore, the politics of a majoritarian issue often relegate social science to the periphery of the policy arena. That fact partly explains why research is often powerless to restrain legislators from enacting multimillion-dollar class-size reduction policies. The California and Indiana cases are representative of the many untargeted class-size policies that have proliferated in the past decade.

Lesson 2: When legislators enact untargeted class-size policies, careful evaluation of the policy is usually an afterthought and requires the use of methodological tools that produce ambiguous findings. When research is commissioned to study the impact of a class-size policy that includes all schools, evaluation suffers. For example, the most comprehensive evaluation of Indiana's policy occurred nearly twenty years after the original Prime Time policy was implemented. Because the Prime Time policy reduced pupil-teacher ratios in all public school districts, researchers conducted multivariate analyses of observational data that revealed ambiguous correlations between small classes and student outcomes rather than clear causal links. By 2002, Indiana's policy was a deeply popular policy resting on little rigorous empirical evidence.

The familiar pattern of research following policy expansion has been replicated in other states with untargeted class-size policies. For example, Nevada spent more than \$500 million during the 1990s to cap class sizes at twenty students in the early grades. Despite the impressive amount of money devoted to reducing pupil-teacher ratios, state legislators have not funded an evaluation of the program. Instead, local districts have paid for limited evaluations, and most studies have revealed little consistent positive effect on student achievement. Therefore, a 1997 report recommended that the state legislature "fund a comprehensive evaluation of the Class-Size Reduction Program," but to date, Nevada legislators have not done so.⁶² In 2002, Florida voters passed an amendment to cap class sizes from pre-kindergarten to twelfth grade by 2010, leaving little opportunity to conduct a rigorous evaluation of the policy. Although some edu-

cators and politicians have used research to encourage a more targeted policy, one sponsor of the Florida amendment dismissed the idea: "People can twist it. They can turn it upside down. But the voters want smaller class sizes one way or other."⁶³ Untargeted class-size policies may be evaluated at some later date, but they yield unclear findings that usually support only one confident conclusion—the need for better research. Yet that recommendation assumes that good research would be used to determine the scope and size of a public policy and that elected officials would be willing to tame the growth of a popular policy.

Lesson 3: Social science research may play a more direct role in helping legislators design targeted class-size policies where the "clients" of government are viewed as legitimate beneficiaries. James Q. Wilson points out that, in making decisions, politicians and their constituencies "take into account not only who benefits but whether it is *legitimate* for that group to benefit."⁶⁴ In *The Black-White Test Score Gap*, Christopher Jencks and Meredith Phillips contend that compensatory educational policies for young children are politically popular precisely because "the beneficiaries appear so deserving. Hardly anyone blames first graders' limited vocabulary on defects of character or lack of ambition. First graders of every race seem eager to please."⁶⁵ However, targeted class-size policies do not benefit everyone; they concentrate benefits among a subgroup of minority and low-income children and high-poverty schools and districts. Although proponents of such policies often invoke normative claims about the value of young children, they cannot assume that political support will form a reliable and broad foundation for legislative action.

The political predicament facing targeted class-size policies opens the door for research to enter the policy arena. Research is welcome because it is needed to garner political support. A targeted class-size reduction policy such as Tennessee's Project Challenge and Wisconsin's SAGE must rely more heavily on empirical evidence as a source of legitimacy because the benefits are concentrated among a small subgroup of disadvantaged children—that is, clients who do not vote, have no money, and often attend racially and economically segregated schools. Good research, then, may increase the odds that a targeted policy will become more popular if there is proof of a clear benefit to clients receiving some form of preferential treatment. In that case, empirical findings can have political consequences: if an evaluation indicates that the treatment had larger effects on the achievement of lower-performing students, then there is evidence that the policy works best for the clients most in need. Statistical interactions should be heeded and main effects ignored. Evidence from Tennessee's STAR experiment most strongly supported a targeted policy, which is precisely what state leaders adopted when they approved Project Challenge.

Although targeted class-size reduction policies are admittedly rare, they do exist and often rely on evidence as a source of legitimacy. For example, in Virginia, the Fairfax County Public Schools, the nation's twelfth-largest school system, has continued to implement a targeted policy based on evaluations of its reduced pupil-teacher ratio initiative, which reduced class sizes to 15 students in first grade in fifteen schools with the highest concentration of low-income and low-achieving students. A 1997 local evaluation concluded that the "Reduced-Ratio Program is more effective for students from low socioeconomic backgrounds." Based on that finding, the authors recommended that resources "be focused in a smaller number of schools in the poorest communities."⁶⁶ In the past decade, the district has continued to allocate money for limited class-size reductions and other policy innovations in twenty low-performing schools that represent less than 10 percent of the district's elementary schools.⁶⁷ Results from local evaluations helped justify the targeted class-size policy, much the same way that research influenced the Tennessee and Wisconsin policies.

Lesson 4: Although different incentives govern the work of scholars and politicians, political decisions can directly influence the quality of social science research. Lawmakers work to satisfy their constituencies, and social scientists work to publish papers in peer-reviewed journals. Ostensibly, there is a clear division between the world of ideas and the world of action. However, the case studies suggest that political decisions can directly affect the methodological quality of social science research. Evaluations of targeted class-size policies may yield more publishable findings, because they usually meet a minimum requirement of quantitative policy analysis—the presence of a control group.⁶⁸ When class reduction is targeted, the treatment is withheld from some students and schools, and it is possible to create a well-matched control group formed through random assignment or a rigorous matching procedure. The availability of an untreated control group is a prerequisite of all evaluation designs that attempt to infer a link between smaller class sizes and student outcomes.

This is one reason why many social scientists—in particular, economists, statisticians, and quantitative sociologists—have exploited data generated by the Tennessee and Wisconsin studies to conduct a number of creative secondary analyses.⁶⁹ Scholars have reaped enormous benefits from these two state policies, published dozens of empirical and theoretical papers based on data, and engaged in open and honest debate about the implications of the findings for research and policy.⁷⁰ By passing legislation to evaluate targeted class-size policies, lawmakers in Tennessee and Wisconsin contributed to the work of social scientists. The reverse was true in California and Indiana, where

the massive scale-up of class-size policies created major hurdles in evaluating the effects of policy on student achievement. The case studies suggest that there is a reciprocal relationship between research and policy: just as social science can inform education policy, so political decisions can shape the rigor and quality of research.

Lesson 5: Once in the policy arena, social scientists do not provide definitive solutions to complex problems like the racial achievement gap; rather, they assume a critical perspective that encourages evaluation of educational policies in terms of their effects on student achievement. Mary Jo Bane has argued that social scientists should be viewed less as “expert problem solvers” than as “participants in democratic deliberation . . . making contributions to public discussions in which we do not control either the outcomes or the use of our findings.”⁷¹ Regardless of their disciplinary training and political perspectives, social scientists share a method—a way of thinking about problems—that performs a valuable function in the policy arena by helping to clarify questions and narrow policy options. They are trained to frame questions that can be addressed with appropriate methods; to test theories that can be falsified with data; to use reliable albeit imperfect measures of complex constructs; to interpret and report results accurately; to highlight alternative explanations of the findings; and never to claim 100 percent confidence about any finding—except the need for more and better research. The goal of social science research is not to prove a point or to rationalize a political decision. Rather, the social science perspective represents a kind of countervailing pressure that may help to restrain policy, focus its goals, and encourage measurement of policy impacts.

When research findings accord with the political views of a large number of citizens and elected officials, they have greater relevance and currency in the policy arena. That is one reason why the results from the Tennessee STAR project have influenced a number of state and district policies. According to economist Gary Burtless, “even though the trend toward smaller class size was under way for decades before Tennessee experimentally tested the impact of class-size reductions, the STAR experiment results were among the first to be widely cited in the popular press as persuasive evidence that smaller classes could be helpful.”⁷² Given the current national effort to “leave no child behind,” policies that improve the achievement of minority children are likely to capture the attention of scholars, legislators, and lay audiences. Better collaborations between politicians and professors might improve the capacity of state government to fund and support experimental evaluations that produce clear findings to guide efforts to address the achievement gap.⁷³

Future Prospects

That targeted class-size reduction policies are seemingly so rare may deepen pessimism about the role that research can play in shaping education policy. But there are also hopeful signs that compensatory education policies reserved primarily for young children from minority and low-income families are politically feasible and scientifically defensible. As clients of government, young children are viewed as legitimate and worthy recipients of additional resources and preferential treatment. The goal of narrowing achievement disparities is etched into the No Child Left Behind Act. Spending money to reduce class sizes for minority students at the beginning of their schooling careers represents a narrowly tailored public policy that serves compelling interests in U.S. society—most notably, the national effort to address achievement disparities. Ideally, intellectual and moral commitment to this task should motivate scholars to encourage good research and effective policies that enhance the quality of education for the most vulnerable members of our democracy.

Notes

1. International evidence also shows little significant relationship between class size and achievement in the middle grades. See, for example, the paper in this volume by Ludger Wößmann. See also, "Special Issue—Class Size: Issues and New Findings," *Educational Evaluation and Policy* (Summer 1999). More recent reviews of the empirical research on class size also are provided by Alan B. Krueger, "Economic Considerations and Class Size," *Economic Journal* 113, no. 485 (2003): 34–63; Eric A. Hanushek, "The Failure of Input-Based Schooling Policies," *Economic Journal* 113, no. 485 (2003): 64–98.

2. Quoted in Lawrence Mishel and Richard Rothstein, eds., *The Class Size Debate* (Washington: Economic Policy Institute, 2002), p. 3.

3. *Ibid.*

4. Criticisms of across-the-board class-size reduction policies and analyses of costs can be found in Eric Hanushek, "Evidence, Politics, and the Class Size Debate" in *The Class Size Debate*, edited by Mishel and Rothstein; John Folger, "Lessons from Class Size Policy and Research," *Peabody Journal of Education* 67, no. 1 (Fall 1989); Dominic Brewer and others, "Estimating the Cost of National Class Size Reductions under Different Policy Alternatives," *Educational Evaluation and Policy Analysis* 21, no. 2 (Summer 1999): 179–92; Caroline Hoxby, "The Effects of Class Size on Student Achievement: New Evidence from Population Variation," *Quarterly Journal of Economics* 115, no. 4 (2000): 1239–285.

5. In a 2006 poll conducted by Public Agenda, 85 percent of superintendents and principals indicated that they would vote for a local school board candidate who believed that "if the public schools finally got more money and smaller classes, they could do a better job." See Public Agenda, *Reality Check 2006* (New York, 2006), p. 20. In addition, 86 percent of new teachers believed that a "very effective" way to improve quality is to "reduce class size." See Public Agenda, *Where We Are Now: Twelve Things You Need to Know about Public Opinion and Public Schools* (New York, 2003), p. 31. In a 1999 survey involving a random sample of 1,422 adults nationwide, a majority

of parents (61 percent) and nonparents (64 percent) thought that overcrowded classrooms were a "major" problem facing the local public schools. When given a list of policies to improve their local schools, 75 percent of parents and 66 percent of nonparents "strongly" favored "reducing class sizes" even if the changes "cost more money and required additional tax dollars." See poll data for NPR/Kaiser/Kennedy School Education Survey (www.npr.org/programs/specials/poll/education/education.results.html 1999).

6. James Q. Wilson, *American Government, Institutions, and Policies*, 3rd ed. (Lexington, Mass.: D. C. Heath, 1986), p. 431.

7. *Ibid.*, p. 432.

8. H. Hugh Hecllo, "Values Underpinning Poverty Programs for Children," *Future of Children* 7 (Summer/Fall 1997): 141-48.

9. James S. Coleman and others, *Equality of Educational Opportunity* (Washington: U. S. Government Printing Office), p. 312.

10. Christopher Jencks, "The Coleman Report and the Conventional Wisdom," in *On Equality of Educational Opportunity*, edited by Frederick Mosteller and Daniel P. Moynihan (New York: Random House, 1972), p. 98.

11. John P. Gilbert and Frederick Mosteller, "The Urgent Need for Experimentation," in *On Equality of Educational Opportunity*, edited by Mosteller and Moynihan, p. 375.

12. Gene Glass and Mary L. Smith, *Meta-Analysis of the Relationship of Class Size and Student Achievement* (San Francisco: Far West Laboratory for Educational Research, 1978); Gene Glass and Mary L. Smith, "Meta-Analysis of Research on Relationship of Class Size and Achievement," *Educational Evaluation and Policy Analysis* 1, no. 1 (1978): 2-16.

13. Glass and Smith, *Meta-Analysis of the Relationship of Class Size and Student Achievement*, p. ii.

14. *Ibid.*

15. Educational Research Services, *Class Size Research: A Critique of Recent Analyses* (Arlington, Va.: 1980).

16. Robert Slavin, "Class Size and Student Achievement: Is Smaller Better?" *Contemporary Education* 62 (Fall 1990): 6-12.

17. Hans J. Eysenck, "An Exercise in Mega-Silliness," *American Psychologist* 33, no. 5 (1978): 517.

18. Gene Glass, "On Criticisms of Our Class Size/Student Achievement Research: No Points Conceded," *Phi Delta Kappan* 62, no. 4 (1980): 244.

19. Henry J. Aaron, *Politics and the Professors* (Brookings, 1978).

20. According to a review of state class-size policies, the Education Commission of the States found that a total of six class-size reduction policies were implemented from 1977 to 1988. From 1989 to 2005, there were a total of twenty-three class size reduction initiatives. See Kyle Zinth, "State Class-Size Reduction Measures" (Denver: Education Commission of the States, March 2005).

21. Gary W. Ritter and Robert F. Boruch, "The Political and Institutional Origins of a Randomized Controlled Trial on Elementary School Class Size: Tennessee's Project STAR," *Educational Evaluation and Policy Analysis* 21 (Summer 1999): 115-16.

22. "On May 22, 1985, House Bill 544 (HB0544) to implement Project STAR was passed in the Tennessee House of Representatives by a vote of 94 to 1; the Tennessee Senate passed the bill on the following day by a vote of 27 to 0." Cited in Ritter and Boruch, "The Political and Institutional Origins of a Randomized Controlled Trial on Elementary School Class Size," p. 117.

23. *Ibid.*, p. 123.

24. According to Frederick Mosteller, the Project STAR study is "one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools." Frederick Mosteller, "The Tennessee Study

of Class Size in the Early Grades," *Future of Children*, vol. 5, *Critical Issues for Children and Youths* (Summer/Fall 1995): 113–26, quote on p. 113.

25. The most relevant findings are reported in publications by Charles Achilles, Jeremy Finn, Eric Hanushek, Alan Krueger, and Frederick Mosteller. See Mosteller, "The Tennessee Study of Class Size in the Early Grades"; Jeremy D. Finn and Charles M. Achilles, "Answers and Questions about Class Size: A Statewide Experiment," *American Educational Research Journal* 27, no. 3 (1990): 557–77; Eric A. Hanushek, "Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects," *Educational Evaluation and Policy Analysis* 21, no. 2 (1999): 143–64; Alan B. Krueger, "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics* 114, no. 2 (1999): 497–532; Alan B. Krueger and Diane M. Whitmore, "The Effect of Attending a Small Class in the Early Grades on College Test Taking and Middle School Test Results: Evidence from Project STAR," *Economic Journal* 111, no. 468 (2001): 1–28; Jeremy D. Finn, Susan B. Gerber, and Jayne Boyd-Zaharias, "Small Classes in the Early Grades, Academic Achievement, and Graduating from High School," *Journal of Educational Psychology* 97, no.2 (2005): 214–23.

26. For example, see the paper in this volume by Diane Whitmore Schanzenbach.

27. It is important to note that Governor Ned McWherter, a Democrat, and his appointed state education commissioner, Charles Smith supported the Project Challenge policy. Despite a change in political leadership from the Alexander to McWherter administration, the class size policy continued the legacy of STAR by using its findings to design a targeted class-size policy. Randy Kenner, "Teachers Given Smaller Classes Call Students Big Winners," *Knoxville News-Sentinel*, November, 11, 1990, p. A1.

28. Personal correspondence from Charles M. Achilles, professor emeritus, School of Education, Eastern Michigan University and Seton Hall University, August 10, 2006.

29. Charles M. Achilles, Jayne B. Zaharias, and Barbara Nye, "Policy Use of Research Results: Tennessee's Project Challenge," paper presented at the annual meeting of the American Educational Research Association, San Francisco (April 1995).

30. Frederick Mosteller, "How Does Class Size Relate to Achievement in Schools?" in *Earning and Learning*, edited by Susan E. Mayer and Paul E. Peterson (Brookings, 1999), pp. 117–30.

31. Charles M. Achilles, Barbara A. Nye, and Jayne Boyd-Zaharias, *Policy Use of Research Results: Tennessee's Project Challenge*, paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1995.

32. Frederick Mosteller, "How Does Class Size Relate to Achievement in Schools?" p. 123.

33. *Ibid.*, p. 18.

34. Jennifer McGiverin, David Gilman, and Chris Tillitski, "A Meta-Analysis of the Relation between Class Size and Achievement," *Elementary School Journal* 90, no. 1 (1989): 47–56.

35. Ritter and Boruch, "The Political and Institutional Origins of a Randomized Controlled Trial on Elementary School Class Size," p. 115.

36. *Ibid.*, pp. 18–29.

37. "Early Grades: Actions to Reduce Class Size," *Education Week*, October 16, 1985.

38. Christopher Tillitski, "The Longitudinal Effect Size of Prime Time, Indiana's State-Sponsored Reduced Class Size Program," *Contemporary Education* 62, no. 1 (1990): 26.

39. Personal correspondence from Daniel K. Lapsley, professor of psychology, Department of Psychology, Notre Dame University, August 1, 2006.

40. Daniel K. Lapsley and others, *Instructional Assistants, Class Size, and Academic Achievement: An Evaluation of Indiana's Prime Time*, report submitted to the Indiana Department of Education (February 2002), p. vi.

41. *Ibid.*, p. 60.

42. Personal correspondence from Jayma Ferguson, Division of Prime Time/Reading First, Indiana Department of Education, November 13, 2006.

43. Marshall S. Smith, "Assessment Trends in a Contemporary Policy Context," in *Analytic Issues in the Assessment of Student Achievement*, edited by David W. Grissmer and J. Michael Ross (U. S. Department of Education, National Center for Education Statistics, 2000), p. 273.

44. Alex Molnar and others, "Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Reduction in Wisconsin," *Educational Evaluation and Policy Analysis* 21 (Summer 1999): 165-78. The SAGE program was initially open to schools in which at least 30 percent of students were from families below the poverty line and from districts where at least 50 percent of students were below the poverty line. Moreover, SAGE consisted of four interventions, and in exchange for \$2,000, districts were required to reduce the student-teacher ratio to 15 students per teacher, beginning with kindergarten and first grade in 1996-97; establish longer school days; develop a rigorous curriculum; and create staff development programs. However, the focus of SAGE became reducing class sizes because most schools claimed to have "rigorous curriculum" in place, longer school days, and staff development. For more details, see John A. Zahorik, Alex Molnar, and Philip Smith, *SAGE Advice: Research on Teaching in Reduced-Size Classes* (Educational Policy Studies Laboratory, College of Education, Arizona State University, 2003).

45. As in any quasi-experimental study, there are possible selection biases that threaten the internal validity of the treatment effects. In the SAGE studies, a number of comparison schools decided to opt out of the study due to the testing burdens and limited benefits of participation. Since 1996-97, seven of the original seventeen comparison schools left the study, and five of those seven applied for and became SAGE schools. According to Thomas Hruz, *The Costs and Benefits of Smaller Classes in Wisconsin* (Thiensville, Wis.: Wisconsin Policy Research Institute, 2000), p. 21: "This problem with the departure of comparison schools relates more generally to concerns of selection bias, which occurs when tested members of a study volunteer to participate because they have a vested interest in the program. The problem is particularly acute in SAGE, where schools apply to participate (and are publicly eager to do so) and where comparisons are made against schools that are forced, seemingly with contempt, into participating."

46. Alex Molnar and others, "Evaluating the SAGE Program," p. 166.

47. *Ibid.*, p. 177. A more recent summary of SAGE results is in Phil Smith, Alex Molnar, and John Zahorik, "Class-Size Reduction, A Fresh Look at the Data," *Educational Leadership* 61, no. 1 (2003): 72-74.

48. Hruz, *The Costs and Benefits of Smaller Classes in Wisconsin*, p. 7.

49. *Ibid.*, p. 34.

50. Norman L. Webb and others, *Participation in the Student Achievement Guarantee in Education (SAGE) Program and Performance on State Assessments at Grade 3 and Grade 4 for Three Cohorts of Students: Grade 1 Students in 1996-97, 1997-98, and 1998-99*, (Madison, Wis.: Wisconsin Center for Education Research, 2004).

51. The three evaluations were carried out by researchers at the University of Wisconsin-Milwaukee, the University of Wisconsin-Madison, and the Wisconsin Policy Research Institute.

52. David C. Illig, "Reducing Class Size: A Review of the Literature and Options for Consideration" (Sacramento, Calif.: California Research Bureau, 1996). The California Research Bureau is akin to the federal Government Accountability Office. Its mission is to provide "nonpartisan research services to the Governor and his staff, to both houses of the legislature, and to other state elected officials."

53. *Ibid.*, p. 2.

54. Thomas B. Timar, "The 'New Accountability' and School Governance in California," *Peabody Journal of Education* 78, no. 4 (2003): 189.

55. George W. Borhnstedt and Brian M. Stecher, eds., *What We Have Learned about Class Size Reduction in California* (Sacramento, Calif.: California Department of Education, 2002), p. 18.

56. Mixed results of the California policy on student outcomes are in Borhnstedt and Stecher, eds., *What We Have Learned about Class Size Reduction in California*. However, based on a

propensity score analysis, recent empirical research suggests that the California policy had a positive impact on math achievement. See Fatih Unlu, "California Class Size Reduction Reform: New Findings from the NAEP," Working Paper (Princeton University, Department of Economics, November 2005).

57. Borhnstedt and Stecher, eds., *What We Have Learned about Class Size Reduction in California*, p. 5.

58. *Ibid.*, p. 6.

59. *Ibid.*, p. 66.

60. For more on the political and institutional origins of the 1996 class size reduction policy, see Borhnstedt and Stecher, eds., *What We Have Learned about Class Size Reduction in California*, chapter 1.

61. Carol H. Weiss, "The Many Meanings of Research Utilization," *Public Administration Review* 39, no. 5 (1979): 427-62. Mary Snow, *An Evaluation of the Class Size Reduction Program* (Carson City, Nev.: Nevada Department of Education, 1997), p. 6.

62. Mary Snow, *An Evaluation of the Class Size Reduction Program* (Carson City, Nev.: Nevada Department of Education, 1997), p. 6.

63. Stephen Hegarty, "On Class Size, Question Is How," *St. Petersburg Times*, December 25, 2002, p. 1B.

64. Wilson, *American Government, Institutions, and Policies*, p. 428.

65. Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (Brookings, 1998), p. 47.

66. James DiStefano, *Evaluation of the Reduced-Ratio Program Final Report* (Fairfax, Va.: Fairfax County Public Schools, Office of Program Evaluation, 1997), p. 30.

67. Michelle P. Serafin, *Project Excel, Final Report: Fall 1999-Spring 2003* (Fairfax County, Va.: Fairfax County Public Schools, Office of Program Evaluation, 2003), p. B-3.

68. Even more sophisticated methods, such as the use of propensity scores to match treatment and control cases, require the treatment to be withheld from some group of students. The logistic regression models used to create propensity scores are based on a binary outcome, which usually denotes whether or not the subject received the treatment. Absent a control group, there is no way to apply propensity score analyses to data. See Paul R. Rosenbaum and Donald B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, no. 1 (1983): 41-55. For a recent extension of propensity score analysis to nonbinary treatments, see Kosuke Imai and David van Dyk, "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score," *Journal of the American Statistical Association* 99, no. 467 (2004): 854-66.

69. For example, see the paper in this volume by Diane Whitmore Schanzenbach.

70. Psychologists and economists interested in studying how the quality of student-teacher interactions is related to improved social and cognitive outcomes have studied the mechanisms inside small classes to help explain improvements in student achievement in the early grades. For relevant research, see National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network, "Does Class Size in First Grade Relate to Children's Academic and Social Performance or Observed Classroom Processes?" *Developmental Psychology* 40, no. 5 (2004): 651-64; Edward Lazear, "Education Production," *Quarterly Journal of Economics* 116, no. 3 (2001): 777-803; Jeremy D. Finn, Gina M. Pannozzo, and Charles M. Achilles, "The 'Why's' of Class Size: Student Behavior in Small Classes," *Review of Educational Research* 73 (Fall 2003): 321-68; Ronald F. Ferguson, "Can Schools Narrow the Black-White Test Score Gap?" in *The Black-White Test Score Gap*, edited by Jencks and Phillips. In "Can Schools Narrow the Black-White Test Score Gap?" (p. 368), economist Ronald Ferguson contends that the "interaction of class size and race is not an especially surprising finding. . . . If, as the evidence indicates, black children are more sensitive than whites to teachers' perceptions, and black children's work habits

and behavioral problems present greater challenges to teachers, smaller classes that are easier for teachers to manage may have more impact on improving black students' scores than whites'. It is probably correct that having fewer pupils per class can improve learning outcomes and reduce racial disparities."

71. Mary Jo Bane, "Presidential Address—Expertise, Advocacy, and Deliberation: Lessons from Welfare Reform," *Journal of Policy Analysis and Management* 20, no. 2 (2001): 191–97. Quote is from p. 195.

72. Gary Burtless, "Randomized Field Trials for Policy Evaluation: Why Not in Education?" in *Evidence Matters*, edited by Frederick Mosteller and Robert Boruch (Brookings, 2002), p. 196.

73. For a recent analysis of the role of nonpartisan research organizations on state policymaking, see John A. Hird, *Power, Knowledge, and Politics* (Georgetown University Press, 2005). According to Hird's analysis, very few state governments have the capacity to carry out complex analytical studies of public policy. With respect to education policy, some state governments (for example, Texas and Florida) have collaborated with university researchers to set up sophisticated longitudinal databases. However, planned experiments, such as those in Tennessee and Wisconsin, are also needed to obtain better data on the causal effects of education policies such as class-size reduction.

Copyright of Brookings Papers on Education Policy is the property of Brookings Institution Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.